Published in final edited form as:

Environ Res. 2019 November ; 178: 108601. doi:10.1016/j.envres.2019.108601.

A Bayesian ensemble approach to combine PM_{2.5} estimates from statistical models using satellite imagery and numerical model simulation

Nancy L. Murray^{a,*}, Heather A. Holmes^b, Yang Liu^c, Howard H. Chang^a

^aEmory University, Department of Biostatistics and Bioinformatics, Atlanta, GA, 30322, USA

^bUniversity of Nevada, Reno, Department of Physics, Reno, NV, 89557, USA

^cEmory University, Department of Environmental Health, Atlanta, GA, 30322, USA

Abstract

Ambient fine particulate matter less than 2.5 μm in aerodynamic diameter (PM_{2.5}) has been linked to various adverse health outcomes. PM_{2.5} arises from both natural and anthropogenic sources, and PM_{2.5} concentrations can vary over space and time. However, the sparsity of existing air quality monitors greatly restricts the spatial-temporal coverage of PM_{2.5} measurements, potentially limiting the accuracy of PM_{2.5}-related health studies. Various methods exist to address these limitations by supplementing air quality monitoring measurements with additional data. We develop a method to combine PM_{2.5} estimated from satellite-retrieved aerosol optical depth (AOD) and chemical transport model (CTM) simulations using statistical models. While most previous methods utilize AOD or CTM separately, we aim to leverage advantages offered by both data sources in terms of resolution and coverage using Bayesian ensemble averaging. Our approach differs from previous ensemble approaches in its ability to not only incorporate uncertainties in PM_{2.5} estimates from individual models but also to provide uncertainties for the resulting ensemble estimates. In an application of estimating daily PM_{2.5} in the Southeastern US, the ensemble approach outperforms previously developed spatial-temporal statistical models that use either AOD or bias-corrected CTM simulations in cross-validation (CV) analyses. More specifically, in spatially clustered CV experiments, the ensemble approach reduced the AOD-only and CTM-only model's root mean squared error (RMSE) by at least 13%. Similar improvements were seen in R². The enhanced prediction performance that the ensemble technique provides at fine-scale spatial resolution, as well as the availability of prediction uncertainty, can be further used in health effect analyses of air pollution exposure.

*Corresponding author. nancy.murray@emory.edu (N.L. Murray).

CRediT authorship contribution statement

Nancy L. Murray: Conceptualization, Methodology, Software, Formal analysis, Writing – original draft. **Heather A. Holmes:** Data curation, Writing - review & editing. **Yang Liu:** Data curation, Writing - review & editing. **Howard H. Chang:** Conceptualization, Methodology, Software, Writing - original draft, Supervision, Funding acquisition.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.envres.2019.108601>.

Keywords

Air pollution; Exposure assessment; Health impact; Spatial modeling

1. Introduction

Air pollution negatively impacts human health, as supported by various studies around the world (Brunekreef and Holgate, 2002; Hoek et al., 2013; Liu et al., 2013; Clark et al., 2014; Evans et al., 2014; Brook et al., 2017). While air pollution represents a complex mixture of chemicals, particulate matter (PM) less than $2.5\ \mu\text{m}$ in aerodynamic diameter ($\text{PM}_{2.5}$), in particular, has received increasing interest in public health (Pui et al., 2014; Hart et al., 2015; Maji et al., 2017). $\text{PM}_{2.5}$ is a mixture of solids and liquids that can penetrate deep into the lower respiratory system to affect the lungs and circulatory system (Brook et al., 2002; Maté et al., 2010; Adam et al., 2015). $\text{PM}_{2.5}$ is composed of primary and secondary PM, with primary PM coming from sources like wildfires, erosion, and pollen; and secondary PM resulting from chemical reactions in the atmosphere (United States Environmental Protection Agency, 2009). Also, sources of $\text{PM}_{2.5}$ include power generation, industrial operations, and automobiles. These anthropogenic emissions and the changing climate can have notable impacts on $\text{PM}_{2.5}$ concentrations and, subsequently, on human health. As a result, the United States Environmental Protection Agency (USEPA) regulates $\text{PM}_{2.5}$ as one of its criteria pollutants to protect public health (Hubbell et al., 2009).

Population-based studies of air pollution and health have contributed significantly to setting regulatory standards worldwide. However, these studies often suffer from routine estimation of exposures using regulatory monitoring networks. Monitors in these networks are preferentially located in specific geographic areas, often in areas with high pollution levels and large populations. Due to high cost of maintenance, $\text{PM}_{2.5}$ monitor measurements are spatially sparse, such that extrapolating these measurements over a large spatial domain may be inappropriate, and are sometimes temporally available only in 1-in-3 or 1-in-6 days time periods.

More recently, an important research area involves developing data fusion products that supplement monitoring measurements with numerical model simulations and remotely sensed observations. These data fusion models typically involve retrospectively estimating exposures. The overarching goal of data fusion is to increase the spatial-temporal coverage of air quality data to support health analyses and health impacts assessments, as seen in existing applications of data fusion products for epidemiological studies involving exposure estimation (Gray et al., 2014; Warren et al., 2016; Chang et al., 2011).

Numerical model simulations used in air pollution research are known as chemical transport models (CTM). CTMs are 3-dimensional deterministic models that simulate gridded air pollution concentrations based on state-of-the-art knowledge on drivers of air quality (Chipperfield, 1999). Advantages of CTM include its complete spatial-temporal coverage and the ability to incorporate chemical and physical processes associated with air pollution. However, CTM is computationally expensive and, as a result, is often only available at crude

spatial resolutions. Because CTMs are often archived and shared, CTMs continue to be used in estimating $PM_{2.5}$.

Remotely sensed aerosol optical depth (AOD) has been examined extensively in its ability to predict $PM_{2.5}$ in combination with other meteorological and land use variables (Liu et al., 2005, 2009). AOD measures the degree to which aerosols prevent light from penetrating the atmosphere. AOD measurements can come from both polar orbiting or geostationary satellites (Levy et al., 2007; Zhou et al., 2018). We focus on polar orbiting satellites in this work. Some main advantages of satellite-based AOD are its fine spatial resolution, global coverage, and public accessibility. However, remotely sensed data represents columnar measurements and can suffer from missing data due to retrieval error and cloud cover.

CTM simulations and AOD values cannot be used directly in health analyses because complex spatial-temporal bias exists when compared to ground-level monitoring data (Marmur et al., 2006; Friberg et al., 2017, 2018; Loría-Salazar et al., 2017). For example, the Community Multiscale Air Quality (CMAQ) model, a type of CTM, may suffer from underprediction or overprediction due to error in inputs and discretization over space and time (Mebust et al., 2003; Lim et al., 2010). AOD measures aerosol over the entire atmospheric column and its relationship with ground-level $PM_{2.5}$ can depend on various factors. Therefore, statistical data fusion models that calibrate CTM and AOD data against observed measurements are needed (Berrocal et al., 2010; Chang et al., 2014).

Most existing data fusion models that incorporate uncertainty quantification have been developed to utilize only one data source: CTM or satellite AOD. Concurrent utilization of both data sources in the fusion process may provide more accurate $PM_{2.5}$ estimates. Specifically, CTM simulation can address the missing data problem in satellite AOD, while satellite AOD can provide additional fine-scale spatial information to CTM simulation. Current approaches center around using CTM simulations to impute missing AOD values, followed by using the gapfilled AOD field as a predictor of $PM_{2.5}$ in multi-stage regression models (de Hoogh et al., 2018; Xiao et al., 2017) or machine learning algorithms (Di et al., 2016; Hu et al., 2017; Reid et al., 2015). (Kloog et al., 2015) provides full coverage by first fitting a model with available data, then smoothing predictions from this model to achieve complete spatial-temporal coverage. Because of this multi-stage approach, obtaining prediction standard error is challenging. Similarly, in the Global Burden of Disease project of Van Donkelaar et al., annual $PM_{2.5}$ averages are obtained by using satellite AOD values that are informed by CTM simulations to account for the vertical aerosol profile (Van Donkelaar et al., 2016). However, one cannot conduct an epidemiological study that examines short-term effects of air pollution based on results from the Global Burden of Disease project (Brauer et al., 2012). Most of the aforementioned studies rely on long-term averages, which do not capture the complex daily missingness of AOD data. Our method uses a sophisticated statistical model to estimate daily $PM_{2.5}$ exposure while also propagating uncertainty.

In this article, we describe a way to combine estimates of $PM_{2.5}$ from spatial-temporal statistical models using Bayesian ensemble averaging. Specifically, predictions from statistical data fusion models using either CTM simulation or satellite AOD are combined

with spatially varying weights. The focus on statistical models is motivated by the need to provide uncertainties in $PM_{2.5}$ estimates, in terms of prediction standard error, that can be used in subsequent health effect and health impact analyses. Our model-based approach offers several advantages compared to previous methods, namely the ability to incorporate various sources of uncertainty in predictions and to characterize the relative prediction performance of CTM versus satellite AOD. Model-based approaches also provide data-driven information on relationships between $PM_{2.5}$ and predictors that is often not available from algorithmic, or machine learning, methods. In an application, we evaluate the proposed method for predicting daily $PM_{2.5}$ in the Southeastern United States (Southeastern US) using 12 km CTM simulations and 1 km satellite-derived AOD.

2. Methods

2.1. Data

We obtained daily ground-level 24-h average measurements of $PM_{2.5}$ from 63 monitors in the Southeastern US over the period 2003 to 2005 via the USEPA's Air Quality System (AQS). We strategically use this period of time in order to perform subsequent health analyses with data from the same time period. CTM simulations were obtained from the USEPA Models-3/Community Multiscale Air Quality (CMAQ) model version 4.5 at a 12 km $\times$ 12 km horizontal spatial resolution (Byun and Schere, 2006). We acquired satellite-retrieved AOD measurements by the aerosol remote sensor Moderate Resolution Imaging Spectroradiometer (MODIS), which orbits the Earth on the National Aeronautics and Space Administration's Aqua and Terra satellites. We utilized a new multiangle implementation of atmospheric correction (MAIAC) algorithm that provides AOD values at a 1 km $\times$ 1 km spatial resolution (Lyapustin et al., 2011a, b). For each AOD grid cell, we also compiled variables including: elevation from the US Geological Survey, forest cover and road lengths from the 2001 National Land Cover data, meteorology (e.g. wind speed) from the North American Land Data Assimilation Systems, and $PM_{2.5}$ primary emission point sources from the 2002 USEPA National Emissions Inventory. As in Hu et al., forest cover and elevation were averaged from their original resolutions of about 1 km and about 30 m, respectively, to the 1 km $\times$ 1 km MAIAC grid cell level (Hu et al., 2013). Additionally, road lengths and point emissions were summed over the 1 km $\times$ 1 km MAIAC grid cell level.

Fig. 1 shows the locations of the 63 AQS monitors in our study region and gridded $PM_{2.5}$ simulations from CMAQ for an example day of March 17, 2005. Similarly, Fig. 2, with an overlay of the same AQS monitor locations, shows the 1 km-level satellite MAIAC AOD values on the same day with a considerable amount of missing data. Overall, the MAIAC AOD is missing for approximately 57% of the days and grid cells in our study. The differences in spatial resolution are also apparent between CMAQ and MAIAC AOD in Figs. 1 and 2.

2.2. Statistical modeling

2.2.1. Bayesian hierarchical modeling for daily $PM_{2.5}$ —We first describe the model for combining monitoring data with CMAQ outputs or AOD retrievals as predictors for point-referenced AQS monitoring measurements in a Bayesian spatial-temporal

hierarchical model (BHM). Predictions of PM_{2.5} from the PM_{2.5}-CMAQ BHM and PM_{2.5}-AOD BHM are subsequently used as inputs to the ensemble model.

Let $Y(s, t)$ represent the observed PM_{2.5} concentration on day t at locations s . Following Berrocal et al. and Chang et al., our statistical model has the form of a BHM:

$$Y(s, t) = \alpha_1(s) + \alpha_2(s)X(s, t) + \beta_1(t) + \beta_2(t)X(s, t) + \mathbf{Z}(s, t)\boldsymbol{\gamma} + \varepsilon(s, t), \quad (1)$$

where $X(s, t)$ is the linked AOD or CMAQ values in the grid cells containing the monitor at locations s . $\mathbf{Z}(s, t)$ is a vector of additional predictors with coefficient $\boldsymbol{\gamma}$ (Berrocal et al., 2010; Chang et al., 2014). For the AOD model, $\mathbf{Z}(s, t)$ includes the following land use and meteorology variables: elevation, forest cover, road length, primary emission source, wind speed, and temperature. Because CMAQ uses information on emissions and meteorology to perform simulations, $\mathbf{Z}(s, t)$ is not included in the PM_{2.5}-CMAQ BHM. Preliminary analysis also showed that including additional covariates does not improve prediction performance for the CMAQ model. Finally, the residual error term, $\varepsilon(s, t)$, is independent normally distributed with mean zero and variance σ_ε^2 .

Parameters $\alpha_1(s)$ and $\alpha_2(s)$ in Equation (1) are the spatial random intercept and spatial random slope, respectively; $\beta_1(t)$ and $\beta_2(t)$ in Equation (1) are the temporal random intercept and temporal random slope, respectively. $\alpha_1(s)$, $\alpha_2(s)$, $\beta_1(t)$, and $\beta_2(t)$ are sometimes referred to as calibration parameters because they correct for the additive and multiplicative bias associated with CMAQ or AOD. Additional details about the modeling assumptions for BHM can be found in the Supporting Information.

2.2.2. Combining estimates from statistical models—Our proposed method to combine PM_{2.5} estimates from the CMAQ-only and AOD-only model is based on the Bayesian Model Averaging (BMA) framework. BMA has been applied to probabilistic weather forecasting in order to combine forecasts from different numerical weather models (Raftery et al., 2005). Here, we extend the approach for estimating spatial-temporal air pollution concentrations when predictions from multiple statistical models are available. To our knowledge, this framework has not previously been used in modeling spatial-temporal air pollution.

We consider the following model:

$$p(y_{st} | M_1, M_2) = w_s f_1(y_{st} | M_1) + (1 - w_s) f_2(y_{st} | M_2), \quad (2)$$

Where y_{st} is the PM_{2.5} value; $f_k(y_{st} | M_k)$ is the posterior *predictive* distribution of y_{st} from model M_k , and w_s is the weight for the PM_{2.5}-CMAQ BHM at location s . w_s ranges from 0 to 1, with a default value of 1 on days where AOD is missing.

Equation (2) can be viewed as a predictive model, where w_s is the posterior probability (ensemble weight) that the PM_{2.5}-CMAQ BHM is the better estimate of PM_{2.5} at monitor s . Here we assume $f_k(y_{st} | M_k) \equiv \phi(y_{st} | \mu_{st}^{(k)}, \sigma_{st}^{2, (k)})$, i.e. a Normal posterior predictive distribution of y_{st} with mean $\mu_{st}^{(k)}$ and variance $\sigma_{st}^{2, (k)}$ using either the PM_{2.5}-CMAQ BHM (k

= 1) or the PM_{2.5}-AOD BHM ($k=2$). Hence, the point predictions of y_{st} can be defined by its posterior mean

$$\hat{y}_{st} = w_s \mu_{st}^{(1)} + (1 - w_s) \mu_{st}^{(2)} \quad (3)$$

which is a weighted average of predictions from the PM_{2.5}-CMAQ BHM and the PM_{2.5}-AOD BHM. Similarly, the error for y_{st} is defined as

$$\hat{\sigma}_{y_{st}}^2 = w_s^2 \sigma_{st}^{2,(1)} + (1 - w_s)^2 \sigma_{st}^{2,(2)} \quad (4)$$

which allows us to quantitatively define uncertainties and make inferences. Bayesian inference also allows us to capture the uncertainty in the weight estimation procedure. Additionally, the posterior interval can be defined as the 2.5% and the 97.5% interval of the conditional distribution. To allow for spatial interpolation of the ensemble weight to locations without monitors, we further assume that $q_s = \text{logit}(w_s)$ is a Gaussian process with an exponential covariance function, i.e. $\text{Cov}(q_s, q_{s'}) = \tau^2 e^{-\|s-s'\|/\rho}$, where τ^2 controls the spatial variability, and ρ controls the rate of spatial decay in dependence.

2.2.3. Estimation and prediction—Estimation and prediction are accomplished in seven stages, which we describe in the enumerated steps below.

1. Fit the PM_{2.5}-CMAQ BHM using Equation (1) to obtain posterior predictive means, $\mu_{st}^{(1)}$, and variances, $\sigma_{st}^{2,(1)}$, for each day and location.
2. Fit the PM_{2.5}-AOD BHM using Equation (1) to obtain posterior predictive means, $\mu_{st}^{(2)}$, and variances $\sigma_{st}^{2,(2)}$, for each day and location where we have observed AOD values.
3. Create out-of-sample CMAQ-based predictions. Randomly leave 10% of the PM_{2.5} observations out then obtain prediction means and prediction variances using the remaining 90% of the data. Repeat this ten times. Stack the predictions to create a dataset.
4. Create out-of-sample AOD-based predictions. Randomly leave 10% of the PM_{2.5} observations out then obtain prediction means and prediction variances using the remaining 90% of the data. Repeat this ten times. Stack the predictions to create a dataset. Note: the training folds and validation folds are the same for CMAQ and AOD.
5. Estimate spatially varying weights based on PM_{2.5} measurements and out-of-sample prediction datasets from Steps 3 and 4 using Equation (2).
6. Interpolate the weights to 1 km × 1 km grid cells using kriging.
7. Combine the estimates from Steps 1 and 2 using weights from Step 5 in the same fashion as Equation (3) to obtain the *ensemble estimate*.

Notice in Steps 3 and 4, to avoid overfitting while estimating ensemble weights, we fit the BHMs repeatedly, but we leave-out and back-predict observations in a *cross-validation*

experiment, similar to approaches employed in stack regression and SuperLearner techniques (LeBlanc and Tibshirani, 1996; Polley and van der Laan, 2010).

Estimation and inference are carried out in a Bayesian framework by specifying priors for all model parameters. Markov chain Monte Carlo (MCMC) methods are used to obtain samples from posterior distributions; we use Gibbs sampler when the full conditional distributions are in closed-form and the random-walk Metropolis-Hasting algorithm otherwise. MCMC computations are standard for Bayesian hierarchical modeling and are provided elsewhere (Chang et al., 2014). MCMC details for fitting the BHM (Step 1 and Step 2) and the ensemble weights (Step 5) are provided in the Supporting Information.

We also investigate alternative approaches to estimate the ensemble weights. In addition to the 10-fold cross-validation (CV) predictions to estimate the weights, we also use a spatial (leave-one-monitor-out) CV approach. We also consider estimating the weights by using a two-stage approach, which first estimates the optimal weight at each model separately, then performs spatial interpolation in a second stage. This method differs from Step 5 in that the uncertainty in the monitor-specific weight is not accounted for in the spatial interpolation. Finally, we compare BHM and the ensemble method to non-Bayesian mixed models to demonstrate the improvements provided by the Bayesian methods.

We use R version 3.5.1 for all estimation and prediction (R Core Team, 2018). The MCMC algorithm is available, coded in R, through the corresponding author's Github site. Sample data is also posted.

2.2.4. Assessing model performance—We evaluated the prediction performance of the proposed ensemble approach using three out-of-sample cross-validation (CV) experiments. First, in a 10-fold CV, we randomly divided the dataset into 10 subsets. Repeatedly, we left out each subset (10% of the data) and used the other 90% of the data to fit the prediction model. Because data are available at each monitor in each CV fold, this 10-fold CV experiment allowed us to evaluate the model's ability to perform *temporal* interpolation when daily PM_{2.5} is missing at monitoring locations.

We also performed a *spatial* CV experiment where all observations at each monitor were left out one-monitor-at-a-time. This allowed us to evaluate the model's ability to perform *spatial* interpolation to estimate PM_{2.5} at locations without monitors.

Finally, we performed *spatially clustered* CV, where 20 clusters formed through hierarchical clustering by proximity of monitoring locations (the *hclust* function in the *stats* package of R) were dropped, and the remaining data were used to estimate PM_{2.5} at multiple locations without monitors (leave one-cluster-at-a-time out) (Johnson, 1967). The twenty clusters, as well as more details about their formation, are given in Fig. S1 in the Supporting Information. The spatially clustered CV simulates a more realistic scenario where the modeling approaches are tasked with spatially interpolating a larger group of spatially missing data rather than a single missing location (Young et al., 2016).

We quantified the performance of different methods using the following statistics: prediction root-mean-square error (RMSE), 95% coverage probability of the posterior intervals (PI),

average posterior standard deviation (SD), and R^2 . R^2 and RMSE were calculated based on posterior predictive means of the left-out observed $PM_{2.5}$ concentrations. Posterior prediction intervals were based on the 2.5th and the 97.5th percentiles of the posterior distribution of the two-component predictive model distribution in Equation (2).

3. Results

Due to the deterministic construction of CMAQ simulations, we have full spatial-temporal coverage for CMAQ across the Southeastern US during the study period of 2003–2005. AOD, on the other hand, is only available at 57.4% of locations and days. $PM_{2.5}$ observations from AQS monitoring sites are available at 75.8% of all days over the three-year study period. Observed $PM_{2.5}$ has a mean of $14.54 \mu\text{g}/\text{m}^3$ and a standard deviation of $7.02 \mu\text{g}/\text{m}^3$. The mean value of $PM_{2.5}$ as determined by the CMAQ simulation is 12% lower at $12.78 \mu\text{g}/\text{m}^3$. Mean AOD is 0.24. Pearson correlations show moderate linear relationships between observed $PM_{2.5}$ and CMAQ at 0.57 and observed $PM_{2.5}$ and AOD at 0.54. CMAQ and AOD are weakly correlated with a Pearson correlation coefficient of 0.13.

Table 1 gives model performance results for the i) 10-fold CV experiment, ii) spatial CV experiment, and iii) spatially clustered CV experiment. Overall, the ensemble approach resulted in improved out-of-sample predictions. Specifically, using inputs derived from the 10-fold CV, the ensemble model achieved the lowest RMSE and highest R^2 in all three evaluations, with the RSME of the ordinary CV being 43% of the standard deviation of $PM_{2.5}$ measurements. The decrease in posterior prediction SD for the 10-fold CV is particularly significant (about 30% reduction), while maintaining the proper coverage of at least 95%. While results of the spatial CV and spatially clustered CV show similar trends as in the 10-fold CV experiment, we find the improvement of the ensemble approach over separate models tends to be better, suggesting the ensemble approach is particularly beneficial for spatial interpolation compared to using only CMAQ or only AOD. Among the three CV experiments, prediction performance decreases from 10-fold CV to spatial CV due to the need to spatially interpolate to locations without monitors; prediction performance also decreases from spatial CV to spatially clustered CV because the number of nearby monitors to aid in interpolation is limited. Despite these differences in CV experiment results, the ensemble approach continues to outperform separate statistical models in all three types of CV experiments. Using a two-stage estimation approach resulted in a negligible reduction in prediction performance compared to the joint estimation method utilized above; these results can be found in the Supporting Information in Table S1 and Table S2. The spatial CV prediction inputs result in weaker results but still justify use of our method compared to the individual models, as seen in the Supporting Information in Table S2. For model performance results of non-Bayesian mixed models and, subsequently, further justification for using the proposed ensemble approach, see the Supporting Information and Table S3. Table 1 can be compared to Table S3, and clear performance advantages from using the Bayesian methods are demonstrated through comparison of each performance metric, i.e. RMSE, coverage, average SD, and R^2 .

Fig. 3 clearly demonstrates the need for spatially varying weights due to the $PM_{2.5}$ -CMAQ BHM receiving a higher assigned weight value for the predictive model in certain areas,

whereas the PM_{2.5}-AOD BHM receives higher weights in more rural areas but also close to some urban centers across the study time period. We can also illustrate spatially kriging the weight estimates from the 10-fold CV experiment to areas without monitoring locations at a finer spatial resolution of 1 km × 1 km across the Southeastern US (Fig. S2).

While the derived weights spatially vary across the Southeastern US, the metropolitan Atlanta, GA area is a clear example of the varying weights the PM_{2.5} CMAQ BHM receives within a relatively small geographical area. PM_{2.5}'s environmental health effects are well-documented in Atlanta, GA (Alhanti et al., 2016; Gass et al., 2015). To that end, we illustrate the use of ensemble estimates of PM_{2.5} within the 20-county metropolitan Atlanta, GA area. We aim to contrast results from the two individual AOD and CMAQ models with our results from the combined, ensemble method. This Atlanta region contains 16,063 AOD grid cells and 143 CMAQ grid cells.

Fig. 4 demonstrates the applicability of the ensemble approach for a single day. The 20-county metropolitan Atlanta area has 9 AQS monitors, but the ensemble approach, combined with spatial kriging and interpolation, allows us to extend the use of weights beyond areas with monitors to obtain posterior predictive mean PM_{2.5} concentrations across a wider swath of land. Fig. 4(AOD model) shows the spatially refined PM_{2.5} estimates from the PM_{2.5}-AOD BHM. Fig. 4(CMAQ model), the PM_{2.5}-CMAQ BHM results, starkly differs from Fig. 4(Ensemble model), the ensemble averaged results, in terms of smoothness. On this particular day, the PM_{2.5}-AOD BHM predicts lower PM_{2.5} concentrations over Atlanta than the PM_{2.5}-CMAQ BHM (Fig. 4(AOD model) and Fig. 4(CMAQ model); also seen in Supporting Information Figs. S3 and S5). The standard error of the PM_{2.5}-AOD BHM is also lower than that of the PM_{2.5}-CMAQ BHM (Figs. S4 and S6). The ensemble approach leads to an average of the PM_{2.5}-AOD BHM and the PM_{2.5}-CMAQ BHM predictions and, thereby, allows for depictions of seamless PM_{2.5} estimates between neighboring spatial fields for which CMAQ alone does not have the complexity.

Fig. 5 displays the long-term 3-year PM_{2.5} concentration estimates over Atlanta from the PM_{2.5}-AOD BHM (Fig. 5(AOD model)), the PM_{2.5}-CMAQ BHM (Fig. 5(CMAQ model)), and ensemble averages restricted to days when AOD was observed (Fig. 5(Ensemble model (AOD observed))) or across all days (Fig. 5(Ensemble model (all days))). The combination of information from the PM_{2.5}-AOD BHM and PM_{2.5}-CMAQ BHM permits more granularity in the maps on both a daily level (Fig. 4(Ensemble model)) and when averaging across days where AOD is observed (Fig. 5(Ensemble model (AOD observed))). This finer resolution on a daily level or on days with observed AOD will aid in acute environmental health effect analyses. However, in Fig. 5(Ensemble model (all days)), the predictions from the PM_{2.5}-CMAQ BHM dominate, likely due to the large amount of temporally missing AOD in this region over the study time period (about 57%).

4. Discussion

Instead of relying solely upon numerical CTM simulations or satellite data to perform data fusion, the proposed combined statistical model framework allows us to incorporate both sources of information and harness their collective predictive power. Existing statistical

methods that combine data, such as Bayesian melding, require modeling of the entire unknown pollution surface, which is computationally intensive, and the data source with the largest sample size (i.e. CTM) will dominate over monitoring measurements and satellite imagery. The proposed ensemble method does not incur significant additional computational burden because it estimates ensemble weights from 10-fold CV predictions, which are routinely performed by researchers when comparing prediction performance of different models.

Another advantage of the ensemble approach entails accounting for differences in spatial resolution between different gridded data because CTM and satellite data are first calibrated to the point-level using monitoring data via Bayesian hierarchical modeling. Finally, in our $PM_{2.5}$ application, the ensemble approach also naturally accounts for the missing values in satellite retrievals, providing $PM_{2.5}$ estimates with complete spatial-temporal coverage. Specifically, in settings with more than two inputs, when satellite AOD is missing, ensemble weights for different inputs can be reweighted among available inputs. In the current version of the ensemble method, missing AOD results in assigning the other input (CTM) a weight of one and proceeding with the estimation. This differs from existing approaches where AOD needs to be imputed before being used as a predictor for $PM_{2.5}$, increasing computational burden and introducing another source of prediction uncertainty.

The computation time of estimating weights for the ensemble method itself is not limiting because it is based on the number of monitors in the area. However, the computation time for predictions do take some time at the fine-scale resolution of $1 \text{ km} \times 1 \text{ km}$. For this reason, we displayed results in Atlanta, GA instead of the entire Southeastern region. Specifically, we presented predictions that incorporated spatial correlation between grid cells (i.e. predicting “maps” of $PM_{2.5}$ concentrations jointly), but if that is not needed, weights can be spatially interpolated one grid cell at a time, which is highly parallelizable.

Our approach deviates from the existing methods in our lack of imputed AOD because of our focus on uncertainty quantification, an important assessment other methods, such as machine learning ensemble methods, cannot provide. Specifically, our Bayesian modeling framework provides prediction standard error, which can then be used for uncertainty quantification in subsequent health effect analyses. Although here we focus on ambient air pollution for the application of this method, the approach is also highly relevant to the estimation of other environmental exposures (e.g. temperature, precipitation) that utilize information from both satellite imagery and numerical model simulations.

This current case study presents a relatively small geographic area; however, the method can be extended to other regions with different meteorological and land characteristics as well. A previous analysis using a non-Bayesian ensemble approach, which does not provide prediction error quantification, to combine estimates from $PM_{2.5}$ -AOD BHM and $PM_{2.5}$ -CMAQ BHM in Colorado has shown similar improvements (Geng et al., 2018).

While this is not the first use of Bayesian Model Averaging (BMA) to perform ensemble modeling in a spatial setting, we focus on combining statistical models rather than deterministic outputs from climate model simulations (Berrocal et al., 2007). Bhat et al. also

use spatial-temporal BMA but to combine global climate projections (Bhat et al., 2011). In contrast, our method interpolates the ensemble weights and uses the model on a much more localized level. Specifically, although the monitoring stations are not randomly placed, we are still able to obtain fine-scale spatially smoothed estimates due to the use of CMAQ and AOD. We also reconcile the spatial resolution differences between our statistical model estimates. Finally, as previously mentioned, we can handle data with high spatial sparsity as demonstrated by the performance of the spatial and spatially clustered CVs.

Several extensions of the proposed method warrant additional investigations. First, ensemble modeling can be generalized to consider multiple sources of information. For example, one can consider a model only driven by fine-scale land use variables with AOD missing. Specifically, the two-component predictive model utilized here can be extended to have multiple weights (i.e. more than two) that are estimated with a multinomial latent variable. In the air pollution application, this may include (1) CTM simulations driven by different assumptions on emission levels and pollution composition for each emission source, (2) multiple satellite parameters that may inform different characteristics of aerosol, and (3) AOD retrievals from different satellites. Our model inputs are also not limited to BHM; we can adapt this method to obtain spatial weights based on the performance of other popular techniques that provide prediction standard error such as kriging or machine learning techniques such as random forests. We modeled spatially varying weights largely due to the ability of satellite-retrieved AOD to predict $PM_{2.5}$ over large areas at a fine scale in certain areas and the error in CMAQ simulation being likely to exhibit spatial variation. Another extension of the ensemble method is to allow weights to depend on spatial and temporal covariates (e.g. land use and meteorology). This may further improve $PM_{2.5}$ prediction and provide insights into which factors are associated with the relative under-performance of the $PM_{2.5}$ -CMAQ BHM and $PM_{2.5}$ -AOD BHM.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This research was supported by the National Institute of Environmental Health Sciences of the National Institutes of Health under award number R01-ES027892. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

- Adam M, Schikowski T, Carsin AE, Cai Y, Jacquemin B, Sanchez M, Vierkötter A, Marcon A, Keidel D, Sugiri D, Al Kanani Z, 2015 Adult lung function and long-term air pollution exposure. ESCAPE: a multicentre cohort study and meta-analysis. *Eur. Respir. J* 45, 38–50. [PubMed: 25193994]
- Alhanti BA, Chang HH, Winquist A, Mulholland JA, Darrow LA, Sarnat SE, 2016 Ambient air pollution and emergency department visits for asthma: a multi-city assessment of effect modification by age. *J. Expo. Sci. Environ. Epidemiol* 26, 180–188. [PubMed: 26350981]
- Berrocal VJ, Raftery AE, Gneiting T, 2007 Combining spatial statistical and ensemble information in probabilistic weather forecasts. *Mon. Weather Rev* 135, 1386–1402.
- Berrocal VJ, Gelfand AE, Holland DM, 2010 A spatio-temporal downscaler for output from numerical models. *J. Agric. Biol. Environ. Stat* 15, 176–197. [PubMed: 21113385]

- Bhat KS, Haran M, Terando A, Keller K, 2011 Climate projections using bayesian model averaging and space–time dependence. *J. Agric. Biol. Environ. Stat* 16, 606–628.
- Brauer M, Amann M, Burnett RT, Cohen A, Dentener F, Ezzati M, Henderson SB, Krzyzanowski M, Martin RV, Van Dingenen R, et al., 2012 Exposure assessment for estimation of the global burden of disease attributable to outdoor air pollution. *Environ. Sci. Technol* 46, 652–660. [PubMed: 22148428]
- Brook RD, Brook JR, Urch B, Vincent R, Rajagopalan S, Silverman F, 2002 Inhalation of fine particulate air pollution and ozone causes acute arterial vasoconstriction in healthy adults. *Circulation* 105, 1534–1536. [PubMed: 11927516]
- Brook RD, Newby DE, Rajagopalan S, 2017 The global threat of outdoor ambient air pollution to cardiovascular health: time for intervention. *JAMA Cardiol* 2, 353–354. [PubMed: 28241232]
- Brunekreef B, Holgate ST, 2002 Air pollution and health. *The Lancet* 360, 1233–1242.
- Byun D, Schere KL, 2006 Review of the governing equations, computational algorithms, and other components of the Models-3 Community Multiscale Air Quality (CMAQ) modeling system. *Appl. Mech. Rev* 59, 51–77.
- Chang HH, Reich BJ, Miranda ML, 2011 Time-to-event analysis of fine particle air pollution and preterm birth: results from North Carolina, 2001–2005. *Am. J. Epidemiol* 175, 91–98. [PubMed: 22167746]
- Chang HH, Hu X, Liu Y, 2014 Calibrating MODIS aerosol optical depth for predicting daily PM_{2.5} concentrations via statistical downscaling. *J. Expo. Sci. Environ. Epidemiol* 24, 398–404. [PubMed: 24368510]
- Chipperfield M, 1999 Multiannual simulations with a three-dimensional chemical transport model. *J. Geophys. Res.: Atmosphere* 104, 1781–1805.
- Clark LP, Millet DB, Marshall JD, 2014 National patterns in environmental injustice and inequality: outdoor NO₂ air pollution in the United States. *PLoS One* 9, 1–8.
- de Hoogh K, Hritier H, Stafoggia M, Knzli N, Kloog I, 2018 Modelling daily PM_{2.5} concentrations at high spatio-temporal resolution across Switzerland. *Environ. Pollut* 233, 1147–1154. [PubMed: 29037492]
- Di Q, Kloog I, Koutrakis P, Lyapustin A, Wang Y, Schwartz J, 2016 Assessing PM_{2.5} exposures with high spatiotemporal resolution across the continental United States. *Environ. Sci. Technol* 50, 4712–4721. [PubMed: 27023334]
- Evans KA, Halterman JS, Hopke PK, Fagnano M, Rich DQ, 2014 Increased ultrafine particles and carbon monoxide concentrations are associated with asthma exacerbation among urban children. *Environ. Res* 129, 11–19. [PubMed: 24528997]
- Friberg MD, Kahn RA, Holmes HA, Chang HH, Sarnat SE, Tolbert PE, Russell AG, Mulholland JA, 2017 Daily ambient air pollution metrics for five cities: evaluation of data-fusion-based estimates and uncertainties. *Atmos. Environ* 158, 36–50.
- Friberg MD, Kahn RA, Limbacher JA, Appel KW, Mulholland JA, 2018 Constraining chemical transport pm 2.5 modeling outputs using surface monitor measurements and satellite retrievals: application over the san joaquin valley. *Atmos. Chem. Phys* 18.
- Gass K, Klein M, Sarnat SE, Winquist A, Darrow LA, Flanders WD, Chang HH, Mulholland JA, Tolbert PE, Strickland MJ, 2015 Associations between ambient air pollutant mixtures and pediatric asthma emergency department visits in three cities: a classification and regression tree approach. *Environ. Health* 14, 1–14. [PubMed: 25564290]
- Geng G, Murray NL, Tong D, Fu JS, Hu X, Lee P, Meng X, Chang HH, Liu Y, 2018 Satellite-based daily PM_{2.5} estimates during fire seasons in Colorado. *J. Geophys. Res.: Atmosphere* 123, 8159–8171.
- Gray SC, Edwards SE, Schultz BD, Miranda ML, 2014 Assessing the impact of race, social factors and air pollution on birth outcomes: a population-based study. *Environ. Health* 13, 1–8. [PubMed: 24405644]
- Hart JE, Liao X, Hong B, Puett RC, Yanosky JD, Suh H, Kioumourtoglou MA, Spiegelman D, Laden F, 2015 The association of long-term exposure to PM_{2.5} on all-cause mortality in the Nurses' Health Study and the impact of measurement-error correction. *Environ. Health* 14, 1–9. [PubMed: 25564290]

- Hoek G, Krishnan RM, Beelen R, Peters A, Ostro B, Brunekreef B, Kaufman JD, 2013 Long-term air pollution exposure and cardio-respiratory mortality: a review. *Environ. Health* 12, 1–15. [PubMed: 23286340]
- Hu X, Waller LA, Al-Hamdan MZ, Crosson WL, Estes MG, Estes SM, Quattrochi DA, Sarnat JA, Liu Y, 2013 Estimating ground-level PM_{2.5} concentrations in the southeastern US using geographically weighted regression. *Environ. Res* 121, 1–10. [PubMed: 23219612]
- Hu X, Belle JH, Meng X, Wildani A, Waller LA, Strickland MJ, Liu Y, 2017 Estimating PM_{2.5} concentrations in the conterminous United States using the random forest approach. *Environ. Sci. Technol* 51, 6936–6944. [PubMed: 28534414]
- Hubbell BJ, Crume RV, Evarts DM, Cohen JM, 2009 Policy monitor: regulation and progress under the 1990 Clean Air Act amendments. *Rev. Environ. Econ. Policy* 4, 122–138.
- Johnson SC, 1967 Hierarchical clustering schemes. *Psychometrika* 32, 241–254. [PubMed: 5234703]
- Kloog I, Sorek-Hamer M, Lyapustin A, Coull B, Wang Y, Just AC, Schwartz J, Broday DM, 2015 Estimating daily PM_{2.5} and PM₁₀ across the complex geo-climate region of Israel using MAIAC satellite-based AOD data. *Atmos. Environ* 122, 409–416.
- LeBlanc M, Tibshirani R, 1996 Combining estimates in regression and classification. *J. Am. Stat. Assoc* 91, 1641–1650.
- Levy RC, Remer LA, Dubovik O, 2007 Global aerosol optical properties and application to moderate resolution imaging spectroradiometer aerosol retrieval over land. *J. Geophys. Res.: Atmosphere* 112.
- Lim CY, Stein M, Ching J, Tang R, 2010 Statistical properties of differences between low and high resolution CMAQ runs with matched initial and boundary conditions. *Environ. Model. Softw* 25, 158–169.
- Liu Y, Sarnat JA, Kilaru V, Jacob DJ, Koutrakis P, 2005 Estimating ground-level PM_{2.5} in the eastern United States using satellite remote sensing. *Environ. Sci. Technol* 39, 3269–3278. [PubMed: 15926578]
- Liu Y, Paciorek CJ, Koutrakis P, 2009 Estimating regional spatial and temporal variability of PM_{2.5} concentrations using satellite data, meteorology, and land use information. *Environ. Health Perspect* 117, 886–892. [PubMed: 19590678]
- Liu T, Li TT, Zhang YH, Xu YJ, Lao XQ, Rutherford S, Chu C, Luo Y, Zhu Q, Xu XJ, Xie HY, 2013 The short-term effect of ambient ozone on mortality is modified by temperature in Guangzhou, China. *Atmos. Environ* 76, 59–67.
- Loría-Salazar SM, Panorska A, Arnott WP, Barnard JC, Boehmle JM, Holmes HA, 2017 Toward understanding atmospheric physics impacting the relationship between columnar aerosol optical depth and near-surface PM_{2.5} mass concentrations in Nevada and California, USA, during 2013. *Atmos. Environ* 171, 289–300.
- Lyapustin A, Martonchik J, Wang Y, Laszlo I, Korkin S, 2011a Multiangle implementation of atmospheric correction (MAIAC): 1. Radiative transfer basis and look-up tables. *J. Geophys. Res.: Atmosphere* 116, 1–9.
- Lyapustin A, Wang Y, Laszlo I, Kahn R, Korkin S, Remer L, Levy R, Reid J, 2011b Multiangle implementation of atmospheric correction (MAIAC): 2. Aerosol algorithm. *J. Geophys. Res.: Atmosphere* 116, 1–15.
- Maji KJ, Dikshit AK, Deshpande A, 2017 Disability-adjusted life years and economic cost assessment of the health effects related to PM_{2.5} and PM₁₀ pollution in Mumbai and Delhi, in India from 1991 to 2015. *Environ. Sci. Pollut. Control Ser* 24, 4709–4730.
- Marmur A, Park SK, Mulholland JA, Tolbert PE, Russell AG, 2006 Source apportionment of PM_{2.5} in the southeastern United States using receptor and emissions-based models: conceptual differences and implications for time-series health studies. *Atmos. Environ* 40, 2533–2551.
- Maté T, Guaita R, Pichiule M, Linares C, Díaz J, 2010 Short-term effect of fine particulate matter (PM_{2.5}) on daily mortality due to diseases of the circulatory system in Madrid (Spain). *Sci. Total Environ* 408, 5750–5757. [PubMed: 20825976]
- Mebust MR, Eder BK, Binkowski FS, Roselle SJ, 2003 Models-3 Community Multiscale Air Quality (CMAQ) model aerosol component 2. Model evaluation. *J. Geophys. Res.: Atmosphere* 108, 1–18.

- Polley EC, van der Laan MJ, 2010 Super Learner in Prediction. U.C. Berkeley Division of Biostatistics Working Paper, pp. 1–19.
- Pui DY, Chen SC, Zuo Z, 2014 PM_{2.5} in China: measurements, sources, visibility and health effects, and mitigation. *Particuology* 13, 1–26.
- R Core Team, 2018 R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
- Raftery AE, Gneiting T, Balabdaoui F, Polakowski M, 2005 Using Bayesian model averaging to calibrate forecast ensembles. *Mon. Weather Rev* 133, 1155–1174.
- Reid CE, Jerrett M, Petersen ML, Pfister GG, Morefield PE, Tager IB, Raffuse SM, Balmes JR, 2015 Spatiotemporal prediction of fine particulate matter during the 2008 Northern California wildfires using machine learning. *Environ. Sci. Technol* 49, 3887–3896. [PubMed: 25648639]
- United States Environmental Protection Agency, 2009 Integrated Science Assessment (ISA) for Particulate Matter (Final Report).
- Van Donkelaar A, Martin RV, Brauer M, Hsu NC, Kahn RA, Levy RC, Lyapustin A, Sayer AM, Winker DM, 2016 Global estimates of fine particulate matter using a combined geophysical-statistical method with information from satellites, models, and monitors. *Environ. Sci. Technol* 50, 3762–3772. [PubMed: 26953851]
- Warren JL, Stingone JA, Herring AH, Luben TJ, Fuentes M, Aylsworth AS, Langlois PH, Botto LD, Correa A, Olshan AF, Prevention NBD, 2016 Bayesian multinomial probit modeling of daily windows of susceptibility for maternal PM_{2.5} exposure and congenital heart defects. *Stat. Med* 35, 2786–2801. [PubMed: 26853919]
- Xiao Q, Wang Y, Chang HH, Meng X, Geng G, Lyapustin A, Liu Y, 2017 Full-coverage high-resolution daily PM_{2.5} estimation using MAIAC AOD in the Yangtze River Delta of China. *Remote Sens. Environ* 199, 437–446.
- Young MT, Bechle MJ, Sampson PD, Szpiro AA, Marshall JD, Sheppard L, Kaufman JD, 2016 Satellite-based NO₂ and model validation in a national prediction model based on universal kriging and land-use regression. *Environ. Sci. Technol* 50, 3686–3694. [PubMed: 26927327]
- Zhou M, Laszlo I, Liu H, 2018 Preliminary evaluation of GOES-16 ABI aerosol optical depth product. In: AGU Fall Meeting Abstracts.

Fig. 1. Simulation of $PM_{2.5}$ from the Community Multiscale Air Quality (CMAQ) model at 12 km resolution on March 17, 2005. Black triangles indicate AQS monitoring locations.

Fig. 2.
Satellite-derived aerosol optical depth (AOD) at 1 km $\times$ 1 km gridded resolution on March 17, 2005. Black triangles indicate AQS monitoring locations.

Fig. 3.
Ensemble weights for predictions from the PM_{2.5}-CMAQ Bayesian hierarchical model at AQS monitoring locations.

Fig. 4. Daily estimates of $PM_{2.5}$ concentrations on March 26, 2005 in the 20-county metropolitan Atlanta, GA area using estimates from (top left) the $PM_{2.5}$ -AOD Bayesian hierarchical model (BHM), (top right) the $PM_{2.5}$ -CMAQ BHM, and (bottom left) the ensemble method.

Fig. 5.

Posterior averages of $PM_{2.5}$ concentrations across 2003–2005 in the 20-county metropolitan Atlanta, GA area based on (top left) the $PM_{2.5}$ -AOD Bayesian hierarchical model (BHM), (top right) the $PM_{2.5}$ -CMAQ BHM, (bottom left) the ensemble method for days in the three-year time period where AOD is observed, and (bottom right) the ensemble method for all days in the three-year time period.

Table 1

Prediction performance for daily PM_{2.5} concentrations comparing ensemble averaging with a Bayesian hierarchical model (BHM) using satellite-derived aerosol optical depth (AOD) or a BHM using a numerical model (CMAQ) simulation. Ensemble weights were derived from first performing 10-fold CV.

Evaluation Method	Statistical Model	RMSE	Coverage of 95% PI	Average Posterior SD	R ²
Ordinary (10-fold) CV					
	PM _{2.5} -AOD BHM	3.40	94.07	3.30	0.78
	PM _{2.5} -CMAQ BHM	3.14	95.05	3.28	0.81
	Ensemble	3.00	97.15	2.39	0.83
Spatial (Leave-one-monitor-out) CV					
	PM _{2.5} -AOD BHM	3.45	94.25	3.39	0.77
	PM _{2.5} -CMAQ BHM	3.33	95.32	3.45	0.78
	Ensemble	2.99	96.81	2.38	0.83
Spatially clustered (Leave-one-cluster-out) CV					
	PM _{2.5} -AOD BHM	3.62	94.43	3.59	0.74
	PM _{2.5} -CMAQ BHM	3.93	93.34	3.58	0.69
	Ensemble	3.13	95.73	3.25	0.81

RMSE: root mean squared error (in $\mu\text{g}/\text{m}^3$); PI: prediction interval; SD: standard deviation (in $\mu\text{g}/\text{m}^3$); CV: cross-validation; PM_{2.5}: particulate matter less than 2.5 μm ; AOD: aerosol optical depth; BHM: Bayesian hierarchical model; CMAQ: Community Multiscale Air Quality.